Teacher-Implemented Response Interruption and Redirection: Training, Evaluation, and Descriptive Analysis of Treatment Integrity

Authors:

Aimee Giles, Ph.D., BCBA-D

Email: aimee.giles@southwales.ac.uk

Telephone: +44 (0)1443 482872

Address: School of Psychology, Ferndale 308, University of South Wales, Pontypridd, United Kingdom, CF37 1DL

Shelley Swain, MSc., BCBA

Email: shelleyswain.aba@gmail.com

Telephone: +44 (0)7745581004

Louise Quinn, MSc, BCBA

Email louise.quinn@PBS4.org.uk

Telephone: +44(0)7849581784

Brittany Weifenbach, MSc.

Email: brittanyweifenbach@gmail.com

Telephone: +1 (605) 209-3364

Abstract

Response interruption and redirection (RIRD) is an effective intervention for decreasing stereotypy. During RIRD, contingent on occurrences of stereotypy, therapists interrupt the behavior and prompt the participant to complete an alternative response. Although RIRD has been implemented by teachers in classrooms, it requires continuous monitoring of participants to be implemented with fidelity and may be difficult for teachers to manage. The present study evaluated the effectiveness of RIRD when implemented in classrooms. In addition, we evaluated if novice teaching assistants could be trained to implement RIRD. Finally, a descriptive analysis of treatment integrity errors during RIRD was conducted. Three children and teaching assistants participated. Following a written instructions baseline, the teaching assistants were trained to implement RIRD using modeling, rehearsal, and feedback. The training increased the accuracy of RIRD implementation for all participants. Incorrectly initiating and terminating RIRD were the most common treatment integrity errors observed.

Keywords: Stereotypy, autism spectrum disorder, response interruption and redirection, behavioral skills training, treatment integrity

Biographical statements:

Aimee Giles is a Senior Lecturer of Psychology at the University of South Wales. Her research focuses on staff training, assessment and intervention for challenging behavior, and identifying effective and preferred teaching strategies for individuals with autism spectrum disorder and intellectual disabilities.

Shelley Swain is a Board Certified Behavior Analyst who works with young children and adolescents with autism spectrum disorder and other developmental disabilities.

Louise Quinn is a Board Certified Behavior Analyst working as a Positive Behavior Support Consultant for a private care provider. She provides behavior analytic services to adults with intellectual disabilities and challenging behavior in supported living settings.

Brittany Weifenbach is a teacher at a nonprofit after school program for school age children.

Teacher-Implemented Response Interruption and Redirection: Training, Evaluation, and Descriptive Analysis of Treatment Integrity

Stereotypy is characterized by repetitive motor or vocal behaviors that may be restrictive and resistant to change (Bodfish, Symons, Parker & Lewis, 2000). In addition, stereotypy is more common and more severe in individuals with autism spectrum disorder (ASD) than in individuals with intellectual disabilities (Bodfish et al., 2000). When stereotypy persists for long durations and interferes with day-to-day functioning it may warrant intervention. High levels of stereotypy are often stigmatizing (Jones, Wint, & Ellis, 1990), interfere with social interactions (Wolery, Kirk, & Gast, 1985), and impede skill acquisition (Dunlap, Dyer, & Koegel, 1983; Koegel & Covert, 1972).

High levels of stereotypy in classrooms may disrupt learning for the child engaging in stereotypy and for other children in the classroom (Conroy, Asmus, Sellers & Ladwig, 2005). In these circumstances, practical and effective interventions to decrease stereotypy in classrooms are necessary to ensure learning for all students. However, it may be challenging to identify effective interventions for decreasing stereotypy because this behavior is often maintained by automatic sensory consequences (Cunningham & Schreibman, 2008; Vollmer, 1994). When stereotypy is maintained by automatic reinforcement, it may not be possible to identify and control the exact reinforcer for the behavior (Rapp & Vollmer, 2005). Therefore, function-based interventions may not always be possible and punishment-based interventions may be necessary.

One such intervention that is effective at decreasing rates of stereotypy is response interruption and redirection (RIRD; Ahearn, Clark, MacDonald & Chung, 2007). During RIRD, Ahearn et al. (2007) interrupted occurrences of vocal stereotypy by stating the child's name and initiating eye contact. Next, instructions were delivered that required a vocal response.

Instructions continued to be delivered until the participant complied with three consecutive instructions in the absence of stereotypy. Vocal stereotypy decreased and appropriate vocalizations increased for all participants. In addition, stereotypy remained at low levels for all participants during follow-up.

Substantial decreases in stereotypy following RIRD have been reported across several studies (Ahrens, Lerman, Kodak, Worsdell, & Keegan, 2011; Love, Miguel, Fernand, & LaBrie, 2012; Miguel, Clark, Tereshko, & Ahearn, 2009; Shumacher & Rapp, 2011; Dickman, Bright, Montgomery, & Miguel, 2012). In addition, clients may prefer RIRD to response blocking as an intervention for decreasing stereotypy (Giles, St Peter, Pence, & Gibson, 2012). Although there is substantial evidence to suggest the effectiveness of RIRD for decreasing stereotypy, the majority of these studies have been conducted in controlled, clinical settings (Martinez & Betz, 2013). Therefore, the effectiveness and practicality of RIRD in naturalistic settings like classrooms is still unknown.

To address the efficacy of RIRD implemented in a more naturalistic setting for longer session durations, Martinez, Betz, Liddon, and Werle (2016) evaluated the effectiveness and efficiency of RIRD when implemented during early intensive behavior intervention (EIBI) sessions with one participant with autism. The authors found that stereotypy remained low in sessions with durations of up to 30 min. In addition, the number of instructions per RIRD episode decreased across time. These results provide some promising evidence for using RIRD in less-controlled settings. However, the authors noted that more research is needed because the study was only completed with one participant. In addition, RIRD was implemented by a trained therapist during EIBI sessions. These sessions may be inherently more controlled than other settings (e.g., classrooms). For example, in classrooms, teachers and teaching assistants may

have less experience implementing behavior analytic interventions such as RIRD. They may also be required to supervise multiple students whilst simultaneously delivering 1:1 or small-group instruction. In addition, the instructional tasks in classrooms may be qualitatively different from the types of tasks described by Martinez et al. (2016). Martinez et al. included tasks such as responding to questions and reciprocal statements in addition to tasks the participant has previously mastered. In comparison, instructional tasks in classrooms may include independent work activities (e.g., completing worksheets, independent reading), independent or choral responding during small group instruction, or structured 1:1 teaching.

Liu-Gitz and Banda (2010) evaluated RIRD in a special education classroom with one participant. The intervention was implemented by the child's teacher during ongoing classroom activities when other children were present. Stereotypy decreased relative to baseline. The teacher positively rated the intervention and noted that she had noticed improvements in the participant's behavior. This study provides some additional support for the practicality and effectiveness of RIRD in classrooms. However, the findings are limited because RIRD was only evaluated with one participant. Moreover, it is unclear how the teacher was trained to implement RIRD. Because the majority of studies evaluating RIRD were conducted by trained researchers in controlled settings, it is important to investigate if teachers without formal behavior analytic training can be trained to implement interventions such as RIRD.

Another important implication for implementing RIRD in classrooms is to evaluate the fidelity with which the intervention is implemented. Higher degrees of treatment fidelity are generally associated with better treatment outcomes (e.g., St. Peter Pipkin, Vollmer, & Sloman, 2010; Vollmer, Sloman, & St. Peter Pipkin, 2008). Liu-Gitz and Banda (2010) reported that the teacher in their study delivered RIRD with 100% treatment integrity. However, in busy settings

such as classrooms, it may be difficult to always implement interventions with high fidelity. Teachers are frequently required to monitor multiple students in addition to delivering academic instruction. Under these conditions, it is possible that a teacher may miss occurrences of stereotypy or not deliver RIRD as described by the protocol. At present, it is still unclear what types of treatment integrity errors may commonly occur during classroom implementation of RIRD.

It is important to identify the types of treatment integrity errors that are likely to occur when RIRD is implemented in classrooms so that training efforts may focus on components of RIRD that are more difficult to implement. Furthermore, some types of treatment integrity errors may be more detrimental to intervention effectiveness than others. For example, St. Peter et al. (2010) found that delivery of reinforcers following occurrences of problem behavior was more detrimental to the outcome of a differential reinforcement of alternative behavior intervention than not delivering a reinforcer contingent on the occurrence of the alternative response. Specific to RIRD, Ahrens, Lerman, Kodak, Worsdell, and Keegan (2011) found stereotypy increased when RIRD was delivered on an intermittent schedule. Therefore, RIRD may not be an appropriate intervention for stereotypy in classrooms if it cannot be delivered following each occurrence of stereotypy.

Stereotypy that interferes with learning is a problem in schools that requires an effective and practical intervention that can be implemented with high integrity. Response interruption and redirection is an effective intervention for decreasing stereotypy. However, there is limited research evaluating if RIRD can be implemented with fidelity in naturalistic settings by novice practitioners. Therefore, the purpose of the present study was three-fold. First, we evaluated the effectiveness of RIRD when implemented in classrooms during ongoing activities. Second, we

evaluated if novice teaching assistants could be trained to implement RIRD with their students. Third, we identified the types of treatment integrity errors most commonly made by novice staff trained to implement RIRD.

Method

Participants and Setting

Three children who engaged in motor stereotypy participated in the study. All three children attended a special school for children with ASD. The child participants did not receive any behavior analytic intervention outside of their participation in the study as this was not part of the curriculum or intervention offered by the school. Child participants were identified because stereotypy was disruptive to their learning or the learning of other students in the classroom. Although no formal functional behavior assessment was conducted, information from staff and observations of stereotypy in a variety of settings demonstrated that stereotypy persisted across a number of contexts including small group and 1:1 instruction, independent work periods, and leisure time.

James was a 12-year-old boy diagnosed with ASD. He engaged in motor stereotypy in the form of finger tapping (see Table 1 for operational definitions of stereotypy for all participants). He communicated vocally using complete sentences. Tim was a 9-year-old boy diagnosed with ASD. He engaged in motor stereotypy in the form of hand flapping and body rocking. He communicated using a Picture Exchange Communication System (PECS). Daniel was a 6-year-old boy with ASD. He engaged in motor stereotypy in the form of hand flapping and face tapping. He communicated using PECS. None of the participants had interventions in place for stereotypy.

Three teaching assistants also participated in the study. All three teaching assistants were employed at the special school for children with ASD. Kelly was 26 years old and had been working as a teaching assistant for seven years. Kim was 26 years old and had been working as a teaching assistant for one year. Julie was 33 years old and had been working as a teaching assistant for seven years. All teaching assistants had completed secondary school. They did not report any experience implementing specific interventions for stereotypy. None of the teaching assistants had experience with behavior analytic intervention or RIRD prior to the study.

Staff-child dyads were created based on which teaching assistant worked with each child most frequently. The staff-child dyads were Kelly and James, Julie and Tim, and Kim and Daniel. The dyads remained the same throughout the study. The teaching assistants would occasionally work with other child participants across the course of the day. However, they did not implement the intervention with any child other than the one with which they were paired for the purposes of the study.

All sessions were 5 min in duration and occurred 2-3 times per day 3-4 days per week. Sessions were conducted during ongoing classroom activities during and outside of instructional periods in two different classrooms. Sessions were conducted based on experimenter availability and to ensure a sample of classroom activities were captured. Therefore, sessions were conducted at different points throughout the week across all classroom activities unless the child participants were scheduled for alternate therapies (e.g., speech and language therapy, music therapy, swimming). This schedule was developed to ensure that RIRD was conducted across a full array of classroom activities and that activities were matched across all phases of the study (stereotypy baseline, implementation baseline, and RIRD). Activities that occurred during both

baselines and RIRD included reading, math, math games, writing, computers, art class, and play activities (e.g., painting, puzzles, and tablet games).

Tim and Daniel were students in one class and James was in a different class. Both classrooms contained desks and chairs or tables for the students to work. James' classroom had a computer for student use and a play area with toys. Tim and Daniel's classroom had a seating area with cushions. Some sessions for Tim and Daniel were conducted outside in the playground and in the library. The staff-to-child ratio in both classrooms was one teacher or teaching assistant to three children. During baseline and RIRD sessions, teaching assistants were responsible for supervising 2-3 students (including the child participant) during independent academic work, leisure time, and while assisting the teacher with small-group instruction. During two sessions, the staff-to-child ratio in the classroom was 1:1 due to absences and students being out of the classroom for other activities. One time was during a reading task and one time during a small-group art activity led by the classroom teacher. Across all sessions, at least two or three other children and one or two other teaching assistants not participating in the study were also present in the classroom. In addition, one or two researchers were present with a video camera and a paper for data collection.

Behavioral skills training sessions was conducted in a small, quiet room attached to the classroom which contained a desk, chairs, and a task that would be used during the modeling and rehearsal components of the training. Three researchers were present during the training sessions. Training sessions occurred for an hour and took place across two days.

Response Measurement and Interobserver Agreement

The primary dependent variables in this study were stereotypy and treatment integrity of RIRD implementation. Stereotypy was defined separately for each of the participants and is

detailed in Table 1. Five-minute sessions were divided into 10 s intervals for data collection. Data were collected using momentary time sampling (Tim and Daniel) or partial-interval recording (James). Partial-interval recording was used for James as momentary time sampling consistently underestimated stereotypy when compared to continuous data. During momentary time sampling, behavior was recorded if stereotypy occurred during second 9 and 10 of each interval. During partial-interval recording, behavior was recorded if stereotypy occurred at any point in the interval. Stereotypy was reported as percentage of intervals with stereotypy.

The RIRD procedure was task analyzed to identify the essential components of the intervention based on RIRD procedures described by Giles, St Peter, Pence, and Gibson (2012). Each RIRD component was operationally defined. During RIRD, the teaching assistants were expected to 1) begin RIRD within 3 seconds of stereotypy, 2) use only predetermined instructions, 3) represent the instruction paired with a model prompt if participant did not respond within 3 s, 4) represent the instruction paired with manual guidance if the participant did not respond within 3 s of the model prompt, 5) terminate the RIRD procedure following 3 consecutive responses in the absence of stereotypy, 6) provide neutral praise statement after termination of RIRD procedure, and 7) refrain from delivering reprimands during RIRD.

Treatment integrity of RIRD implementation was measured using an experimenter-developed treatment integrity checklist and converted into a percentage of accurate implementation. Per-opportunity treatment integrity data were collected on accurate implementation of RIRD for the seven components on the checklist. For each opportunity, a component could be recorded as being performed accurately, inaccurately, or not applicable. Components of the RIRD procedure were only scored if the participant initiated RIRD. For example, if the participants did not implement RIRD following an occurrence of stereotypy, only

the component for implementing RIRD within 3 s of stereotypy was scored as inaccurate. Subsequent components were scored as not applicable. Treatment integrity during RIRD was calculated by dividing the total number of correct opportunities by the number of correct plus incorrect opportunities during the session and converted to a percentage.

Following the completion of the study, teaching assistants were provided social validity questionnaire to take away and complete on their own. The questionnaire asked participants to indicate how much they agreed or disagreed with a number of statements. The questionnaire aimed to identify the extent to which the participants found the RIRD intervention easy and useful prior to training, how helpful they found the behavioral skills training, and how they felt implementing the intervention following training. The teaching assistants were informed that their responses on the questionnaire would be anonymous to minimize reactivity.

Interobserver agreement (IOA) was collected for over 33% of sessions from each condition for each of the participants. Agreement data were collected on both stereotypy and treatment integrity by two independent observers via video recordings. For RIRD treatment integrity, agreement coefficients were calculated for accurate and inaccurate responses for each component and averaged to determine overall session IOA. The agreement coefficient was calculated by dividing the smaller number or responses recorded for that component by the larger number recorded for that component. These coefficients were added together and divided by the number of components converted into a percentage. Mean agreement of treatment integrity was 98.5% (range 91-100%) for Kelly, 94.2% (range, 90-98%) for Julie, and 96.8% (range, 92-100%) for Kim. For stereotypy, IOA was calculated by dividing the number of intervals where both observers recorded the occurrence or non-occurrence of stereotypy by the total number of intervals and multiplying by 100. Mean agreement of occurrence of stereotypy was 98.3%

(range, 93-100%) for James, 96.3% (range, 90-100%) for Tim, and 98.2% (range, 90-100%) for Daniel.

Experimental Design

A multiple baseline across participants design was used to evaluate the effects of behavioral skills training on correct implementation of RIRD and stereotypy. The multiple baseline was introduced during the implementation baseline phase.

Procedure

Pre-training. Before any data were collected, observations were conducted in the classrooms to operationally define each child's stereotypy. During the initial observations, instructions that would be later used in RIRD were also assessed to identify instructions that participants could perform independently or that could be manually prompted. James could independently comply with a number of instructions. Because Tim and Daniel were often non-compliant, instructions that were easy to physically prompt were chosen (e.g., "point to the…" or "touch your nose"). Not all instructions were necessarily incompatible with stereotypy (see Table 1 for list of instructions identified for each participant). However, RIRD has been shown to be effective when instructions were not topographically matched to or incompatible with stereotypy (e.g., Ahrens et al., 2011; Giles et al., 2012; Martinez et al., 2016).

Stereotypy Baseline. During the stereotypy baseline, child participants were observed during typical classroom activities and data were collected on stereotypy. Observations were randomly conducted across the school day and occurred during both instructional and non-instructional activities. The teaching assistants were not instructed to interact with children in any specific manner.

Implementation Baseline. During the implementation baseline, the teaching assistants were provided with a flowchart illustrating the instructions on how to implement the RIRD intervention. The flowchart detailed each of the seven components of RIRD. When implementing the RIRD procedure, the teaching assistants were required to say the child participant's name and initiate eye contact contingent on stereotypy. Once the child participant was attending, the teaching assistant provided an instruction from a predetermined list. If the child participant did not respond to an instruction within 3 seconds, or if they responded incorrectly, the instruction was repeated with a model prompt. If after the model prompt the child participant still failed to respond correctly, the instruction was repeated and full physical guidance was delivered. After the child participant complied with the instruction a new instruction was delivered until three instructions (prompted or independent) had been completed without any stereotypy.

The flowchart included the operational definition for the child participant's stereotypy and a list of individualized instructions for that child participant. In addition, the chart explained what to do if the child participant was non-compliant, how and when to prompt, when it was appropriate to deliver praise, and when to terminate RIRD. The teaching assistants were instructed to teach and engage with their child participant as they normally would and to follow the instructions for RIRD to the best of their ability whenever stereotypy occurred. If the teaching assistants had any questions during this condition, they were asked to try their best to complete the intervention. For control purposes, the teaching assistants were instructed to not implement RIRD with students other than their assigned child participant or outside of the experimental sessions. Sessions were 5-min in duration (inclusive of the RIRD procedure).

Behavioral Skills Training. Behavioral skills training was conducted following the implementation baseline using procedure similar to those described by Pence, St. Peter, & Giles (2014). The training was conducted by three graduate students in Applied Behavior Analysis and consisted of four components: instruction, modeling, rehearsal, and feedback. During instruction, teaching assistants were provided with the flowchart detailing the RIRD procedure. Each part of the intervention was explained and the teaching assistants were permitted to ask questions. The experimenters also modeled the correct way to implement each component of the procedure in isolation. Feedback was provided on baseline performance for each teaching assistant.

Following the instruction component, the trainers modeled RIRD. During modeling, trainers demonstrated how to implement RIRD during a 5-min role-play with one trainer acting as the child and another as the teaching assistant. Scripts were developed for the role-play. Average interresponse time for stereotypy was calculated from baseline sessions to determine the frequency of stereotypy during the role-play. In addition, the scripts specified when to comply with instructions during RIRD and when to be non-compliant based on average rates of compliance to instructions during the implementation baseline. Therefore, the trainers were able to model the prompting procedures used during RIRD. Following the role-play, the teaching assistants were provided with an additional opportunity to ask questions about RIRD.

The rehearsal component immediately followed modeling. During rehearsal, teaching assistants were given the opportunity to practice RIRD during 5-min role-plays with an experimenter playing the role of their child participant. Individualized scripts were also developed for rehearsal to determine when to engage in stereotypy and when to comply with instructions during RIRD. Prior to each rehearsal, the teaching assistants were provided with the RIRD flowchart and materials for academic activities that they frequently completed with their

student (e.g., math and writing worksheets). The teaching assistants were told to help their student (i.e., the trainer) to complete the activity and to try to implement RIRD following each instance of stereotypy. In-vivo data were collected by a separate observer using the treatment integrity checklist. Following each rehearsal, feedback was provided to each teaching assistant on the components that they had performed correctly and incorrectly. In addition, the trainers briefly described and modeled how to perform components performed incorrectly during the rehearsal. If treatment integrity was below 90% accuracy during a rehearsal, additional rehearsal and feedback sessions were conducted until two consecutive sessions with 90% accuracy was achieved.

Post-training. Post-training sessions were identical to implementation baseline sessions. If treatment integrity of a session was below 90% accuracy, feedback was provided immediately after the session on the components of RIRD that were implemented incorrectly.

Results

Figure 1 depicts the percentage of intervals with stereotypy and percentage of opportunities treatment integrity results for all participants. During the stereotypy baseline, Tim engaged in moderate levels of stereotypy averaging of 33.3% of intervals containing stereotypy. Daniel engaged in low to moderate levels of stereotypy during the stereotypy baseline. Stereotypy was variable during this phase with a mean of 28.5% of intervals containing stereotypy. Stereotypy was on an increasing trend for James during the stereotypy baseline phase. At the end of the stereotypy baseline phase, James was engaging in stereotypy for approximately 50% of intervals.

During the implementation baseline and post-training sessions, RIRD treatment integrity data were not recorded during sessions where stereotypy did not occur. During the implementation baseline, Julie was performing half of the components of RIRD correctly, but

there was a decreasing trend across the phase. Tim's stereotypy decreased to near zero levels during the implementation baseline. Kim performed approximately 50% of RIRD treatment integrity components correctly during implementation baseline sessions. During this phase, stereotypy for Daniel decreased slightly from baseline, but was variable. Kelly's performance was variable during the implementation baseline, but by the end of the phase, she was performing approximately 50% of RIRD treatment integrity components correctly. During behavioral skills training, all participants met the training criteria of over 90% accuracy for two consecutive sessions within four training sessions. Julie met the training criteria in two sessions, Kim met the training criteria in three sessions, and Kelly met the training criteria in four sessions.

Julie's correct implementation of RIRD decreased following behavioral skills training. However, performance feedback was provided following post-training sessions and RIRD treatment integrity increased to above 90%. There was a slight increase in stereotypy for Tim compared to the implementation baseline, but this decreased across the phase. Kim also had a decrease in correct RIRD implementation at the beginning of the post-training phase. Following two sessions with feedback, correct RIRD implementation increased above 90%. Stereotypy for Daniel also increased slightly at the start of post-training, but decreased across the phase. Across the post-training phase, RIRD treatment integrity for Kelly only fell below 90% for one session. Stereotypy for James was variable and occurred at similar levels to the implementation baseline phase. Stereotypy was on an increasing trend at the end of post-training.

Figure 2 depicts the percentage of errors for all treatment integrity errors made during the study. Total number of errors decreased from 244 during pre-training sessions to 70 during post-training sessions. Before behavioral skills training, the largest proportion of errors occurred

initiating RIRD (21.7% of total errors). In general, participants did not initiate RIRD. However, occasionally they initiated the procedure more than 3 s from the start of stereotypy. Another common error was that participants prematurely terminated RIRD and delivered praise. They did not wait to terminate the intervention following three consecutive responses in the absence of stereotypy (18.8% of total errors). Finally, errors were made during the prompting procedure. Participants would often deliver a model prompt before it was necessary (14.1% of total errors) or failed to deliver a prompt at all (4.5% of total errors). Following behavioral skills training, treatment integrity errors were lower in all components except delivering a physical prompt. The most common errors following training were failing to terminate RIRD following three consecutive responses in the absence of stereotypy (5.4% of total errors) and failing to deliver physical prompts (5.1% of total errors). More errors were made in the physical prompt component following behavioral skills training than before. Participants did not deliver reprimands at any point during the study.

Table 2 depicts the most common errors for each participant before and after behavioral skills training. All three participants had fewer errors implementing RIRD following training compared to before training. Julie emitted the highest number of errors when implementing RIRD both before and after training compared to the other participants. However, errors decreased from 97 errors before behavioral skills training to 40 errors during post-training. The most common error for Julie before and after training was when to terminate RIRD and deliver praise. In addition, before training Julie made errors when delivering a model prompt and after training, made errors when delivering the physical prompt.

Prior to training, Kim made 67 errors while implementing RIRD and made only 19 errors during post-training. Before behavioral skills training, Kim made error across most components

of the intervention. Not beginning RIRD within 3 s of stereotypy occurring was the most frequent error before training, but this error only occurred once following behavioral skills training. Errors when terminating RIRD after three instructions without stereotypy and providing praise both decreased following behavioral skills training. The most common errors after training for Kim occurred during prompting.

Kelly made 80 errors before behavioral skills training and 11 errors during post-training. Kelly's most frequent error before and after training was that she did not initiate RIRD within 3 s of stereotypy. In fact, Kelly often did not initiate RIRD at all when stereotypy occurred. However, this error occurred less frequently following training. Kelly also made errors and did not terminate RIRD following three consecutive occurrences of stereotypy.

The results of the social validity questionnaire were mostly positive. All three participants agreed that the flowchart used to display the instructions on RIRD implementation was helpful. Some participants agreed RIRD was easy to implement and others found it challenging. The participants either agreed or strongly agreed to all statements regarding training, which indicated that behavioral skills training was helpful to improve understanding of the intervention and helped make it easier to implement. However, one participant was neutral on the statement that training improved her confidence to implement RIRD. All participants agreed that RIRD was easier to implement following training. One participant agreed that RIRD was effective while the other two participants were neutral on this statement. Two participants agreed that RIRD was acceptable to use in a classroom and one participant was neutral.

Discussion

Teaching assistants were trained to implement RIRD with students during ongoing classroom activities. In addition, RIRD reduced stereotypy for two of the three child participants.

Overall, treatment integrity errors decreased following behavioral skills training. The most frequent errors following training were not implementing RIRD within 3 s of stereotypy and not correctly using physical prompts during RIRD.

This study adds to the existing literature on RIRD in several ways. First, this study demonstrates a naturalistic evaluation of RIRD in classrooms. The majority of empirical studies on RIRD have been conducted in controlled settings with trained therapists (Martinez & Betz, 2013). In the present study, RIRD was implemented in a special education classroom across the school day during instructional and leisure activities with a staff-to-student ratio of 1:3. For two participants (Tim and Daniel), stereotypy decreased compared to baseline.

Stereotypy is often difficult to treat because there are limited function-based intervention options. There are other options for decreasing stereotypy (e.g., noncontingent reinforcement and differential reinforcement). However, these may not be appropriate for classrooms as they may require providing access to preferred items that may interfere with academic instruction (Lanovaz & Sladeczek, 2012). For these reasons, it may be challenging to identify practical and effective interventions for stereotypy in classrooms. The results of the present study are promising because they demonstrate RIRD is a feasible and effective intervention for use in classrooms. In addition, RIRD was reported as an appropriate intervention for use in classrooms by two of the three teaching assistants on the social validity questionnaire.

Second, this study demonstrated that novice teaching staff could be trained to implement RIRD with high fidelity. One potential barrier to implementing RIRD in naturalistic settings such as classrooms is that teachers are responsible for managing multiple students while simultaneously engaging their students in instruction. Response interruption and redirection requires teachers to constantly monitor and maintain close proximity to students to implement

the intervention each time stereotypy occurs. These requirements may pose practical challenges for teachers implementing RIRD with students. The teaching assistants in the present study were frequently responsible for one or two other students in addition to the child participating in the study.

During the implementation baseline, the teaching assistants implemented RIRD with relatively low treatment integrity. This condition is similar to providing the teachers with a written behavior plan. Performance during the implementation baseline suggests that simply providing a written plan was not sufficient to train the teaching assistants to do RIRD. However, following a 2 hr training, the teaching assistants were able to implement RIRD with high degrees of treatment fidelity. All of the teaching assistants required additional feedback in the classroom following behavioral skills training. Following feedback, RIRD treatment integrity increased to high levels. Some staff may require additional training and support to generalize skills acquired during training to naturalistic settings.

The results of the social validity questionnaire indicated that the teaching assistants found the training on RIRD to be useful and increased their confidence while implementing RIRD. Interestingly, although RIRD effectively decreased stereotypy for two of the three participants, only one teaching rated the intervention as being effective. Because the questionnaires were anonymous, we don't know if the teaching assistant who rated RIRD as being effect was linked to one of the children whose stereotypy decreased during RIRD. When evaluating interventions, it is important to understand stakeholder perceptions of intervention effectiveness to ensure socially valid interventions and behavior change. Although we are unable to further explore the teaching assistant's perceptions of effectiveness in the present study, future research may

investigate the characteristics of interventions and behavior change that are correlated with stakeholder's perceptions of effectiveness.

Third, this study evaluated the types of treatment integrity errors made during RIRD prior to and following behavioral skills training. Prior to training, the most frequent treatment integrity errors were not implementing RIRD within 3 s of stereotypy and not terminating RIRD and delivering praise following three consecutive responses without stereotypy. Following training, fewer treatment integrity errors were made overall, but participants did not always implement RIRD within 3 s of stereotypy and there was an increase in errors during the physical prompt component of RIRD. The increase in this error likely occurred because during the implementation baseline phase, participants did not always initiate RIRD following each occurrence of stereotypy. Therefore, they did not have the opportunity to use the physical prompt. Following training, participants were initiating RIRD more and therefore, there were more opportunities to deliver physical prompts.

In general, the best treatment outcomes occur when interventions are implemented with high fidelity (St. Peter Pipkin et al., 2010; Vollmer et al., 2008). However, the results of the present study are somewhat mixed. For example, Tim's stereotypy decreased during post-training concurrently with increases in Julie's treatment integrity. In contrast, the lowest levels of stereotypy observed for Daniel occurred during the implementation baseline when Kim was implementing approximately 50% of RIRD components correctly. Compared to the implementation baseline, levels of stereotypy were slightly elevated during post-training. On one hand, these results suggest that RIRD may still be effective under less-than-optimal treatment integrity. On the other, it is possible that the effects of RIRD do not maintain over extended periods of time. The majority of evaluations of RIRD have been conducted during 5- or 10-min

evaluations. However, Martinez, Betz, Liddon, and Werl (2016) demonstrated that low levels of stereotypy during RIRD can generalize from controlled treatment rooms to the classroom for up to 30 min observations with one participant. Martinez et al. reported that RIRD became more effective and efficient over the duration of the study. It is possible that similar effects for Daniel may have occurred if we had conducted additional sessions because there was a decreasing trend in stereotypy during post-training. Unfortunately, additional data were not collected because the school year ended.

For all participants, RIRD was only implemented during 5 min sessions embedded across the school day. There were no additional stimuli programmed during RIRD sessions to signal that RIRD would be delivered contingent on stereotypy. However, the presence of a second observer and video camera during sessions may have functioned as a cue that RIRD would be implemented. From the child's perspective, it is possible that RIRD implementation was inconsistent because for much of the school day, RIRD was not delivered contingent on stereotypy. This inconsistency may have affected the overall effectiveness of RIRD during sessions or lower levels of stereotypy may have generalized outside of the session. We did not collect data on stereotypy outside of sessions so it is unclear if the effects of RIRD generalized beyond the 5 min sessions.

Response interruption and redirection consists of two primary components: interruption and redirection. In the present study, participants had difficulty during both interruption and redirection components of RIRD. It is possible that errors during one component of RIRD may differentially affect the overall effectiveness of the intervention. Errors during the interruption component occurred when participants did not implement RIRD at all following an occurrence of stereotypy or when RIRD was delayed by more than 3 s. Research on RIRD and response

interruption suggests that this type of treatment integrity error may detrimentally affect the effectiveness of RIRD. Specifically, Ahrens et al. (2011) found that stereotypy increased when RIRD was not implemented following each instance of stereotypy. Kliebert, Tiger, and Toussaint (2011) found that automatically reinforced behavior (i.e., skin picking and hair twirling) increased when response interruption did not follow each instance of the target response or was delayed by as little as 3 s. Taken together, the results of Ahrens et al. and Kliebert et al. suggest that treatment integrity errors during the interruption component may be particularly problematic.

Treatment integrity errors during the redirection component occurred when participants did not consistently terminate RIRD following three responses (i.e., errors during the redirection component). Saini, Gregory, Uran, and Fantetti (2015) compared RIRD requiring three responses before terminating the procedure to RIRD requiring one response for decreasing stereotypy. One-response RIRD is analogous to terminating RIRD prematurely in the present study. Saini et al. found one-response RIRD was as effective as and more efficient than three-response RIRD. These results suggest that prematurely terminating RIRD may not have a detrimental effect on treatment outcomes. Therefore, errors during the redirection component of RIRD may be less problematic than errors during the interruption component.

Future research should continue to evaluate the critical components that contribute to the effectiveness of RIRD. However, previous research (e.g., Ahrens et al., 2011; Kliebert et al. 2011; & Saini et al. 2015) and the decreased levels of stereotypy during the implementation baseline of the present study may suggest that interrupting stereotypy during RIRD is more critical than redirecting to a different response. Therefore, training on RIRD may focus more specifically on teaching staff the operational definitions for the target behavior and ensuring that

this component is being implemented consistently before training is terminated. If RIRD cannot be implemented consistently following each occurrence of behavior, it may not be the most appropriate or effective intervention.

Some potential limitations of the present study should be noted. First, functional analyses were not conducted to confirm that stereotypy was not sensitive to social consequences. However, we did observe stereotypy to persist across a variety of contexts prior to the start of the study. In addition, because RIRD does not require control over the maintaining reinforcer to be effective, it may still be an appropriate intervention for stereotypy that serves multiple functions (Lui-Gitz & Banda, 2012).

A second limitation was that RIRD was not effective for one participant (James). For James, stereotypy did not decrease even when Kelly was implementing RIRD with near perfect treatment integrity. Because we did not conduct a functional analysis, it is possible that RIRD was not an effective intervention for James because his behavior was sensitive to attention as a reinforcer. James was reported to engage in other behavior to access teacher attention. Although previous research (e.g., Lui-Gitz & Banda, 2010) has suggested RIRD may be appropriate for behavior that is maintained by social and non-social reinforcers, the attention inherently delivered during RIRD may function as a reinforcer for some participants.

A third limitation of the present study is that an error was recorded if RIRD was not implemented within 3 s of stereotypy regardless of if RIRD was eventually delivered or did not occur at all (i.e., error of omission). It is possible that these different types of errors could have very different effects on the outcomes of RIRD. To better understand the critical components of RIRD, it is important for future research to evaluate the different effects of delayed RIRD versus omission errors on stereotypy.

A final limitation is that we conducted relatively brief observations (i.e., 5 min) and did not evaluate the long-term maintenance of the intervention due to the end of the school year. It is unclear if the effects of RIRD would maintain across longer observations or persist over time. Because stereotypy can be disruptive in classrooms, it is likely that interventions like RIRD would need to be implemented across all instructional periods during a school day. Depending on the classroom and child, RIRD may need to be implemented for teaching periods with durations in excess of 10 min. Future research should continue to evaluate the effectiveness and feasibility of RIRD during longer intervention durations.

The maintenance of RIRD treatment integrity is also unclear. Although the teaching assistants were implementing RIRD with high degrees of fidelity during post-training, it is not known if these effects would persist. Decreases in treatment integrity over time could result in increases in stereotypy. However, some research suggests that decreases in treatment integrity may not be as detrimental if there is a recent history of high treatment integrity (St Peter Pipkin et al., 2010). Future research should evaluate the robustness of RIRD during occasional treatment integrity failures.

In addition to evaluating the effects of overall treatment integrity, it is also important for future research to evaluate the effects of specific treatment integrity failures on stereotypy. This information could inform what types of errors have the greatest impact of the effectiveness of the intervention and also the minimum level of treatment integrity necessary for a specific component to maintain the effectiveness of RIRD. To implement RIRD in naturalistic settings such as classrooms, this type of information is important because it can inform how teachers are trained to use RIRD, how frequently treatment integrity should be monitored, and also the

conditions where RIRD is not recommended because it is not feasible to implement key components with sufficient integrity.

References

Ahearn, W. H., Clark, K. M., MacDonald, R. P. F., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 40*, 263–275. doi: 10.1901/jaba.2007.30-06

Ahrens, E. N., Lerman, D. C., Kodak. T., Wordsell, A. S., & Keegan, C. (2011). Further evaluation of response interruption and redirection as treatment for stereotypy. *Journal of Applied Behavior Analysis, 44,* 95-108. doi: 10.1901/jaba.2011.44-95

Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders, 30*, 237-243. doi: 10.1023/A: 1005596502855

Conroy, M. A., Asmus, J. M., Sellers, J. A., & Ladwig, C. N. (2005). The use of an antecedent-based intervention to decrease stereotypic behavior in a general education classroom: A case study. *Focus on Autism and Other Developmental Disabilities, 20*, 223-230. doi: 10.1177/10883576050200040401

Cunningham, A. B. & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders*, *2,* 469-479.

Dickman, S. E., Bright, C. N., Montgomery, D. H., & Miguel, C. F. (2012). The effects of response interruption and redirection (RIRD) and differential reinforcement on vocal stereotypy and appropriate vocalizations. *Behavioral Interventions, 27*, 185-192. doi: 10.1002/bin.1348

Dunlap, G., Dyer, K, & Koegel, R. L. (1983). Autistic self-stimulation and intertrial interval duration. *American Journal on Mental Deficiency, 88*, 194-202.

Giles, A. F., St. Peter, C. C., Pence, S. T., Gibson, A. B. (2012). Preference for blocking or response redirection during stereotypy treatment. *Research in Developmental Disabilities, 33,* 1691-1700. doi: 10.1016/j. ridd.2012.05.008

Jones, R. S. P., Wint, D., & Ellis, N. C. (1990). The social effects of stereotyped behaviour. *Journal of Mental Deficiency Research, 34*, 261-268. doi:10.1111/j.1365-2788.1990.tb01537.x

Kliebert, M. L., Tiger, J. H., & Toussaint, K. A. (2011). An approach to identifying the conditions under which response interruption with reduce automatically reinforced problem behaviour. *Behavior Analysis in Practice, 4(1),* 17-26.

Koegel, R. L., & Covert, A. (1972). The relationship of self-stimulation to learning in autistic children. *Journal of Applied Behaviour Analysis, 5*, 381-387. doi: 10.1901/jaba.1972.5-381

Lanovaz, M. J. & Sladeczek, I. E. (2012). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification, 36*, 146-164. doi: 10.1177/0145445511427192

Liu-Gitz, L. & Banda, D. R. (2010). A replication of the RIRD strategy to decrease vocal stereotypy in a student with autism. *Behavioral Interventions, 25*, 77-87. doi:10.1002/bin.297

Love, J. J., Miguel, C. E, Fernand, J. K., & LaBrie, J. K. (2012). The effects of matched stimulation and response interruption and redirection on vocal stereotypy. *Journal of Applied Behavior Analysis, 45*, 549-564. doi: 10.190 l/jaba.2012.45-549

Martinez, C. K. & Betz, A. M. (2013). Response interruption and redirection: Current research trends and clinical application. *Journal of Applied Behavior Analysis, 46,* 549-554. doi: 10.1002/jaba.38

Martinez, C. K., Betz, A. M., Liddon, C. J., & Werle, R. L. (2016). A progression to transfer RIRD to the natural environment. *Behavioral Interventions, 31,* 144-162. doi: 10.1002/bin.1444

Miguel, C. F., Clark, K., Tereshko, L., & Ahearn, W. H. (2009). The effects of response interruption and redirection and sertraline on vocal stereotypy. *Journal of Applied Behavior Analysis, 42,* 883-888. doi: 10.1901 /jaba.2009.42-883

Pence, S. T., St. Peter, C., & Giles, A. F. (2014). Teacher acquisition of functional analysis methods using pyramidal training. *Journal of Behavioral Education, 23,* 132-149. doi: 10.1007/s10864-013-9182-4

Rapp, J. T., & Vollmer, T. R. (2005) Stereotypy: A review of behavioral assessment and treatment. *Research in Developmental Disabilities, 26,* 527-547. 10.1016/j.ridd.2004.11.005

Saini, V., Gregory, M.. K.,U ran, K. J., & Fantetti, M. A. (2015). Parametric analysis of response interruption and redirection as a treatment for stereotypy. *Journal of Applied Behavior Analysis, 48,* 96-106. doi: 10.1002/jaba.186

Schumacher, B. I., & Rapp, J. T. (2011). Evaluation of the immediate and subsequent effects of response interruption and redirection on vocal stereotypy. *Journal of Applied Behavior Analysis, 44,* 681-685. doi: 10.1901/jaba.2011.44-681

St. Peter Pipkin, C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis, 43,* 47-70. doi: 10.1901/jaba.2010.43-47

Vollmer, T. R. (1994). The concept of automatic reinforcement: Implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities, 15,* 187-207. doi: 10.1016/0891-4222(94)90011-6

Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice, 1,* 4-11.

Wolery, M., Kirk, K., & Gast, D. L. (1985). Stereotypic behavior as a reinforcer: Effects and side effects. *Journal of Autism and Developmental Disorders, 15,* 149-161. doi: 10.1007/BF01531601

Table 1. *Operational definitions of stereotypy and instructions used for each participant.*

Participant	Stereotypy	Instructions used in RIRD
James	Flicking fingers or hands at a rapid pace near the face, while having a tense, extended face and neck and wide eyes. Occasionally extend one arm horizontally, have the other vertical and move with force to either the left or right.	Point to the… Touch the… Clap your hands
Tim	Moving objects rapidly up and down on top of each other, hitting them on the ground or tables. Rocking body back and forth while sitting down.	Clap your hands Touch your nose Touch your tummy Touch your knees Point to the window Point to the floor
Daniel	Clapping one hand on the back of the other, with elbows raised and arms near face. Rapidly moving objects back and forth in front of face, or onto chin.	Clap your hands Touch your nose Touch your tummy Touch your knees Point to the window Point to the floor

Table 2. *Individual participant errors during RIRD.*

	Julie		Kim		Kelly	
Error	Pre training	Post training	Pre training	Post training	Pre training	Post training
Began RIRD within 3 seconds	10 (7.1%)	3 (2.2%)	19 (22.1%)	1 (1.2%)	39 (42.9%)	9 (9.9%)
Used predetermined instruction	1 (0.7%)	0 (0%)	1(1.2%)	1 (1.2%)	0 (0%)	0 (0%)
Terminated after 3 responses without stereotypy	29 (21.2%)	13 (9.5%)	12 (13.9%)	2 (2.3%)	18 (19.8%)	2 (2.2%)
Provided Praise	26 (18.9%)	9 (6.6%)	12 (13.9%)	1 (1.2%)	19 (20.9%)	0 (0%)
Refrained from reprimand	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Model Prompt	27 (19.7%)	6 (4.4%)	13 (15.1%)	7 (8.1%)	4 (4.4%)	0 (0%)
Physical Prompt	4 (2.9%)	9 (6.6%)	10 (11.6%)	7 (8.1%)	0 (0%)	0 (0%)

Note. Numbers in parentheses depict the percentage of total errors made for each participant.

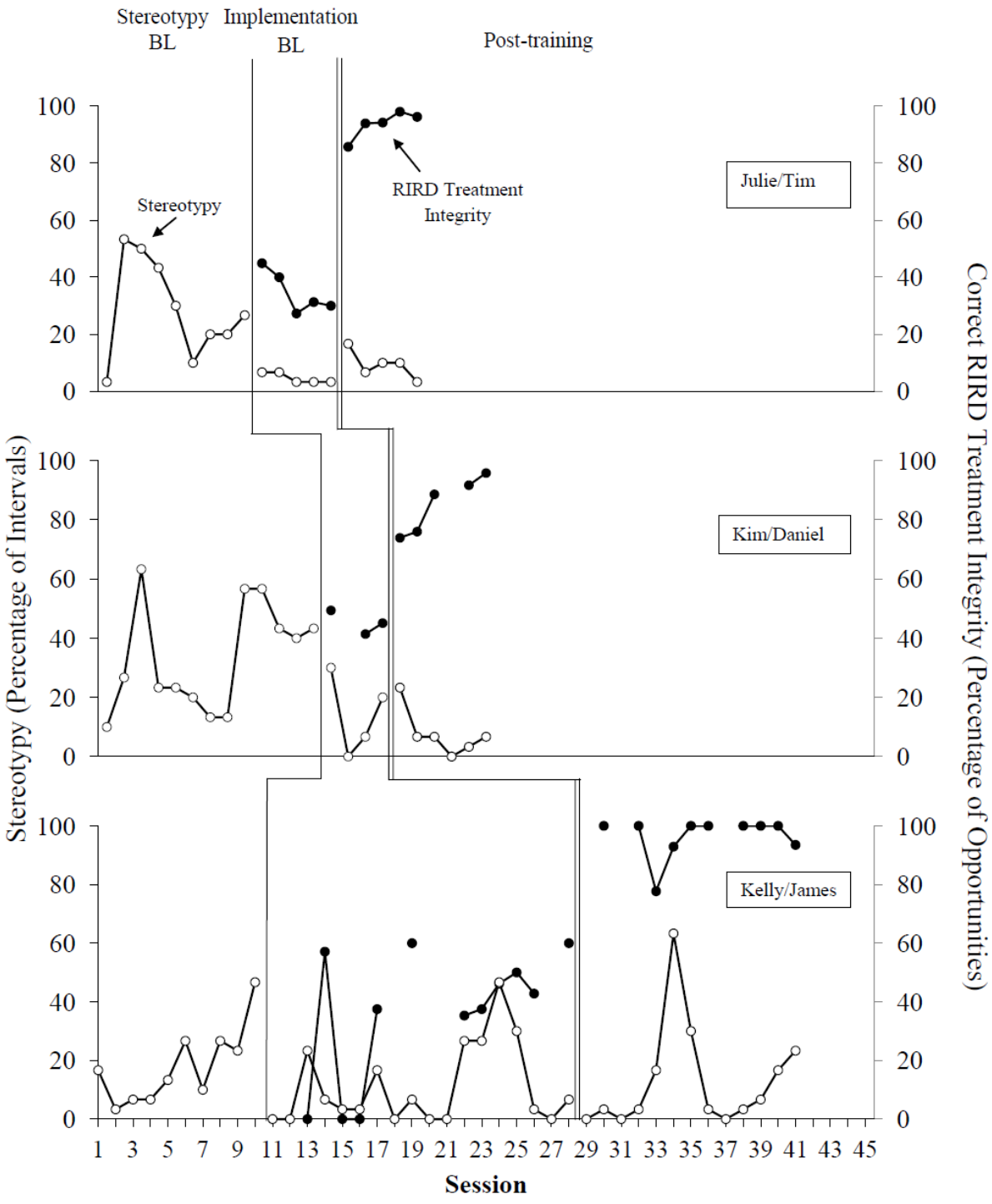

Figure 1. Percentage of intervals with stereotypy and percentage of opportunities correct for RIRD treatment integrity for Julie and Tim (top panel), Kim and Daniel (middle panel), and Kelly and James (bottom panel) across Stereotypy Baseline, Implementation Baseline, and Post-training phases. The double line indicates Behavioral Skills Training (BST). RIRD treatment integrity data were not recorded during sessions where stereotypy did not occur.

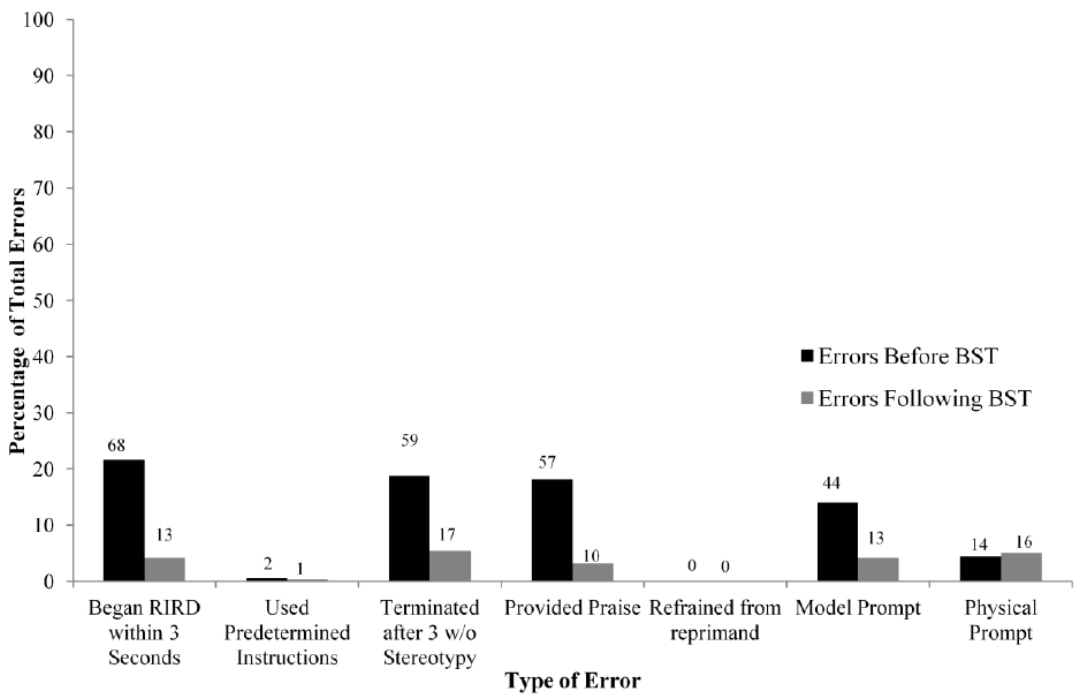

Figure 2. Percentage of total errors before Behavioral Skills Training (BST; black bars) and after BST (gray bars) for all participants. Numbers above each bar indicate the absolute number of errors observed.